C-Feel-It: A Sentiment Analyzer for Micro-blogs

Aditya Joshi[1]   **Balamurali A R**[2]   **Pushpak Bhattacharyya**[1]   **Rajat Mohanty** [3]

[1]Dept. of Computer Science and Engineering, IIT Bombay, Mumbai
[2] IITB-Monash Research Academy, IIT Bombay, Mumbai
[3] AOL India (R&D), Bangalore
India
{adityaj,balamurali,pb}@cse.iitb.ac.in   r.mohanty@teamaol.com

Abstract

Social networking and micro-blogging sites are stores of opinion-bearing content created by human users. We describe *C-Feel-It*, a system which can tap opinion content in posts (called tweets) from the micro-blogging website, Twitter. This web-based system categorizes tweets pertaining to a search string as positive, negative or objective and gives an aggregate sentiment score that represents a sentiment snapshot for a search string. We present a qualitative evaluation of this system based on a human-annotated tweet corpus.

1   Introduction

A major contribution of Web 2.0 is the explosive rise of user-generated content. The content has been a by-product of a class of Internet-based applications that allow users to interact with each other on the web. These applications which are highly accessible and scalable represent a class of media called *social media*. Some of the currently popular social media sites are Facebook (www.facebook.com), Myspace (www.myspace.com), Twitter (www.Twitter.com) etc. User-generated content on the social media represents the views of the users and hence, may be opinion-bearing. Sales and marketing arms of business organizations can leverage on this information to know more about their customer base. In addition, prospective customers of a product/service can get to know what other users have to say about the product/service and make an informed decision.

C-Feel-It is a web-based system which predicts sentiment in micro-blogs on Twitter (called tweets). (Screencast at: http://www.youtube.com/user/cfeelit/) C-Feel-It uses a rule-based system to classify tweets as positive, negative or objective using inputs from four sentiment-based knowledge repositories. A

weighted-majority voting principle is used to predict sentiment of a tweet. An overall sentiment score for the search string is assigned based on the results of predictions for the tweets fetched. This score which is represented as a percentage value gives a live snapshot of the sentiment of users about the topic.

The rest of the paper is organized as follows: Section 2 gives background study of Twitter and related work in the context of sentiment analysis for Twitter. The system architecture is explained in section 3. A qualitative evaluation of our system based on annotated data is described in section 4. Section 5 summarizes the paper and points to future work.

2   Background study

Twitter is a micro-blogging website and ranks second among the present social media websites (Prelovac, 2010). A micro-blog *allows users to exchange small elements of content such as short sentences, individual pages, or video links* (Kaplan and Haenlein, 2010). More about Twitter can be found here [1].

In Twitter, a micro-blogging post is called a *tweet* which can be upto 140 characters in length. Since the length is constrained, the language used in tweets is highly unstructured. Misspellings, slangs, contractions and abbreviations are commonly used in tweets. The following example highlights these problems in a typical tweet:

'*Big brother doing sian massey no favours. Let her ref. She's good at it you know#lifesapitch*'

We choose Twitter as the data source because of the sheer quantity of data generated and its fast reachability across masses. Additionally, Twitter allows information to flow freely and instantaneously unlike FaceBook or MySpace. These aspects of

---

[1]http://support.twitter.com/groups/31-twitter-basics

Twitter makes it a source for getting a live snapshot of the things happenings on the web.

In the context of sentiment classification of tweets Alec et al. (2009a) describes a distant supervision-based approach for sentiment classification. The training data for this purpose is created following a semi-supervised approach that exploits emoticons in tweets. In their successive work, Alec et al. (2009b) additionally use hashtags in tweets to create training data. Topic-dependent clustering is performed on this data and classifiers corresponding to each are modeled. This approach is found to perform better than a single classifier alone.

We believe that the models trained on data created using semi-supervised approaches cannot classify all variants of tweets. Hence, we follow a rule-based approach for predicting sentiment of a tweet. An approach like ours provides a generic way of solving sentiment classification problems in micro-blogs.
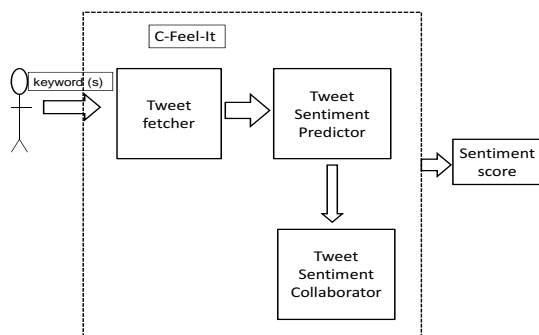
3 Architecture

Figure 1: Overall Architecture

The overall architecture of C-Feel-It is shown in Figure 1. C-Feel-It is divided into three parts: **Tweet Fetcher**, **Tweet Sentiment Predictor** and **Tweet Sentiment Collaborator**. All predictions are positive, negative or objective/neutral. C-Feel-It offers two implementations of a rule-based sentiment prediction system. We refer to them as version 1 and 2. The two versions differ in the Tweet Sentiment Predictor module. This section describes different modules of C-Feel-It and is organized as follows. In subsections 3.1, 3.2 & 3.3, we describe the three

functional blocks of C-FeeL-It. In subsection 3.4, we explain how four lexical resources are mapped to the desired output labels. Finally, subsection 3.5 gives implementation details of C-Feel-It.

Input to C-Feel-It is a *search string* and a *version number*. The versions are described in detail in subsection 3.2.

Output given by C-Feel-It is two-level: tweet-wise prediction and overall prediction. For tweet-wise prediction, sentiment prediction by each of the resources is returned. On the other hand, overall prediction combines sentiment from all tweets to return the percentage of positive, negative and objective content retrieved for the search string.

3.1 Tweet Fetcher

Tweet fetcher obtains tweets pertaining to a search string entered by a user. To do so, we use live feeds from Twitter using an API [2]. The parameters passed to the API ensure that system receives the latest 50 tweets about the keyword in English. This API returns results in XML format which we parse using a Java SAX parser.

3.2 Tweet Sentiment Predictor

Tweet sentiment predictor predicts sentiment for a single tweet. The architecture of Tweet Sentiment Predictor is shown in Figure 2 and can be divided into three fundamental blocks: *Preprocessor, Emoticon-based Sentiment Predictor, Lexicon-based Sentiment Predictor* (refer Figure 3 & 4). The first two blocks are same for both the versions of C-Feel-It. The two versions differ in the working of the Lexicon-based Sentiment Predictor.

Preprocessor

The noisy nature of tweets is a classical challenge that any system working on tweets needs to encounter. Preprocessor deals with obtaining *clean tweets*. We do not deploy any spelling correction module. However, the preprocessor handles extensions and contractions found in tweets as follows.

Handling extensions: Extensions like '*besssssst*' are common in tweets. However, to look up resources, it is essential that these words are normalized to their dictionary equivalent. We replace consecutive occurrences of the same letter (if more than
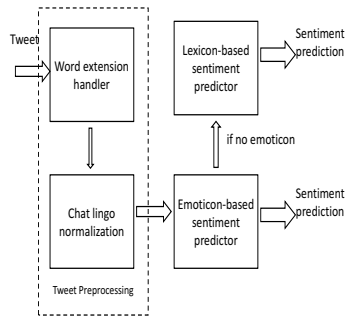
---

[2]http://search.Twitter.com/search.atom

Figure 2: Tweet Sentiment Predictor: Version 1 and 2

Figure 3: Lexicon-based Sentiment Predictor: C-Feel-It Version 1

three occurrences of the same letter) with a single letter and replace the word.

An important issue here is that extensions are in fact strong indicators of sentiment. Hence, we replace an extended word by two occurences of the contracted word. This gives a higher weight to the extended word and retains its contribution to the sentiment of the tweet.

Chat lingo normalization: Words used in chat/Internet language that are common in tweets are not present in the lexical resources. We use a dictionary downloaded from http://chat.reichards.net/ . A chat word is replaced by its dictionary equivalent.

Emoticon-based Sentiment Predictor

Emoticons are visual representations of emotions frequently used in the user-generated content on the Internet. We observe that in most cases, emoticons pinpoint the sentiment of a tweet. We use an emoticon mapping from http://chat.reichards.net/smiley.shtml. An emoticon is mapped to an output label: positive or negative. A tweet containing one of these emoticons that can be mapped to the desired output labels directly. While we understand that this heuristic does not work in case of sarcastic tweets, it does provide a benefit in most cases.

Lexicon-based Sentiment Predictor

For a tweet, the Lexicon-based Sentiment Predictor gives a prediction each for four resources. In addition, it returns one prediction which combines the four predictions by weighting them on the ba-
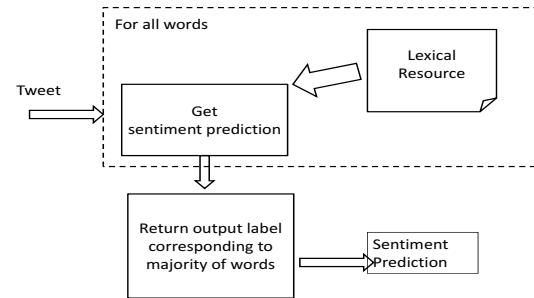
sis of their accuracies. We remove stop words [3] from the tweet and stem the words using Lovins stemmer (Lovins, 1968). Negation in tweets is handled by inverting sentiment of words after a negating word. The words *'no', 'never', 'not'* are considered negating words and a context window of three words after a negative words is considered for inversion. The two versions of C-Feel-It vary in their Lexicon-based Sentiment Predictor. Figure 3 shows the Lexicon-based Sentiment Predictor for version 1. For each word in the tweet, it gets the prediction from a lexical resource. We use the intuition that a positive tweet has positive words outnumbering other words, a negative tweet has negative words outnumbering other words and an objective tweet has objective words outnumbering other words.

Figure 4 shows the Lexicon-based Sentiment Predictor for version 2. As opposed to the earlier version, version 2 gets prediction from the lexical resource for some words in the tweet. This is because certain parts-of-speech have been found to be better indicators of sentiment (Pang and Lee, 2004). A tweet is annotated with parts-of-speech tags and the POS bi-tags (i.e. *a pattern of two consecutive POS*) are marked. The words corresponding to a set of optimal POS bi-tags are retained and only these words used for lookup. The prediction for a tweet uses majority vote-based approach as for version 1. The optimal POS bi-tags have been derived experimentally by using top 10% features on information gain-based-pruning classifier on polarity dataset by (Pang and Lee, 2005). We used Stanford POS tagger(Tou,

---

[3]http://www.ranks.nl/resources/stopwords.html

129

2000) for tagging the tweets.

Note: The dataset we use to find optimal POS bi-tags consists of movie reviews. We understand that POS bi-tags hence derived may not be universal across domains.
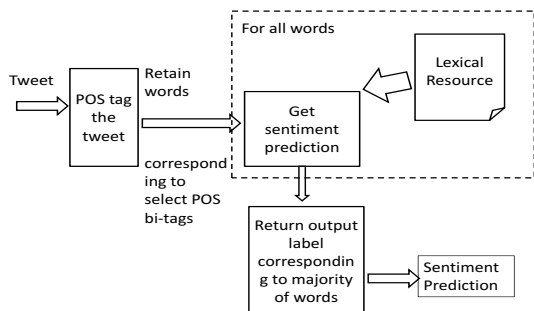

Figure 4: Lexicon-based Sentiment Predictor: C-Feel-It Version 2

3.3 Tweet Sentiment Collaborator

Based on predictions of individual tweets, the Tweet Sentiment Collaborator gives overall prediction with respect to a keyword in form of percentage of positive, negative and objective content. This is on the basis of predictions by each resource by weighting them according to their accuracies. These weights have been assigned to each resource based on experimental results. For each resource, the following scores are determined.

$$posscore[r] = \sum_{i=1}^{m} p_i w_{pi}$$

$$negscore[r] = \sum_{i=1}^{m} n_i w_{ni}$$

$$objscore[r] = \sum_{i=1}^{m} o_i w_{oi}$$

$$where$$

$posscore[r]$ = Positive score for search string r

$negscore[r]$ = Negative score for search string r

$objscore[r]$ = Objective score for search string r

m = Number of resources used for prediction

p_i, n_i, o_i = Positive, negative & objective count of tweet predicted respectively using resource i

w_{pi}, w_{ni}, o_{oi} = Weights for respective classes derived for each resource i

We normalize these scores to get the final positive, negative and objective pertaining to search string r. These scores are represented in form of percentage.

3.4 Resources

Sentiment-based lexical resources annotate words/concepts with polarity. The completeness of these resources individually remains a question. To achieve greater coverage, we use four different sentiment-based lexical resources for C-Feel-It. They are described as follows.

1. SentiWordNet (Esuli and Sebastiani, 2006) assigns three scores to synsets of WordNet: positive score, negative score and objective score. When a word is looked up, the label corresponding to maximum of the three scores is returned. For multiple synsets of a word, the output label returned by majority of the synsets becomes the prediction of the resource.

2. Subjectivity lexicon (Wiebe et al., 2004) is a resource that annotates words with tags like parts-of-speech, prior polarity, magnitude of prior polarity (weak/strong), etc. The prior polarity can be positive, negative or neutral. For prediction using this resource, we use this prior polarity.

3. Inquirer (Stone et al., 1966) is a list of words marked as positive, negative and neutral. We use these labels to use Inquirer resource for our prediction.

4. Taboada (Taboada and Grieve, 2004) is a word-list that gives a count of collocations with positive and negative seed words. A word closer to a positive seed word is predicted to be positive and vice versa.

3.5 Implementation Details

The system is implemented in JSP (JDK 1.6) using Net-Beans IDE 6.9.1. For the purpose of tweet annotation, an internal interface was written in PHP 5 with MySQL 5.0.51a-3ubuntu5.7 for storage.

4 System Analysis

4.1 Evaluation Data

For the purpose of evaluation, a total of 7000 tweets were downloaded by using popular trending topics of 20 domains (like books, movies, electronic gadget, etc.) as keywords for searching tweets. In order to download the tweets, we used the API provided by Twitter [4] that crawls latest tweets pertaining to keywords.

Human annotators assigned to a tweet one out of 4 classes: positive, negative, objective and objective-spam.

---

[4]http://search.twitter.com/search.atom?

A tweet is assigned to objective-spam category if it contains promotional links or incoherent text which was possibly not created by a human user. Apart from these nominal class labels, we also assigned the positive/negative tweets scores ranging from +2 to -2 with +2 being the most positive and -2 being the most negative score respectively. If the tweet belongs to the objective category, a value of zero is assigned as the score.

The spam category has been included in the annotation as a future goal of modeling a spam detection layer prior to the sentiment detection. However, the current version of C-Feel-It does not have a spam detection module and hence for evaluation purpose, we use only the data belonging to classes other than objective-spam.

4.2 Qualitative Analysis

In this section, we perform a qualitative evaluation of actual results returned by C-Feel-It. The errors described in this section are in addition to the errors due to misspellings and informal language. These erroneous results have been obtained from both version 1 and 2. They have been classified into eleven categories and explained henceforth.

4.2.1 Sarcastic Tweets

Tweet: *Hoge, Jaws, and Palantonio are brilliant together talking X's and O's on ESPN right now.*
Label by C-Feel-It: Positive
Label by human annotator: Negative

The sarcasm in the above tweet lies in the use of a positive word '*brilliant*' followed by a rather trivial action of '*talking Xs and Os*'. The positive word leads to the prediction by C-Feel-It where in fact, it is a negative tweet for the human annotator.

4.2.2 Lack of Sense Understanding

Tweet: *If your tooth hurts drink some pain killers and place a warm/hot tea bag like chamomile on your tooth and hold it. it will relieve the pain*
Label by C-Feel-It: Negative

This tweet is objective in nature. The words 'pain', 'killers', etc. in the tweet give an indication to C-Feel-It that the tweet is negative. This misguided implication is because of multiple senses of these words (for example, '*pain*' can also be used in the sentence '*symptoms of the disease are body pain and irritation in the throat*' where it is non-sentiment-bearing). The lack of understanding of word senses and being unable to distinguish between them leads to this error.

4.2.3 Lack of Entity Specificity

Tweet: *Casablanca and a lunch comprising of rice and fish: a good sunday*
Keyword: Casablanca

Label by C-Feel-It: Positive
Label by human annotator: Objective

In the above tweet, the human annotator understood that though the tweet contains the keyword '*Casablanca*', it is not Casablanca about which sentiment is expressed. The system finds a positive word '*good*' and marks the tweet as positive. This error arises because the system cannot find out which sentence/parts of sentence is expressing opinion about the target entity.

4.2.4 Coverage of Resources

Tweet: *I'm done with this bullshit. You're the psycho not me.*
Label by SentiWordNet: Negative
Label by Taboada/Inquirer: Objective
Label by human annotator: Negative

On manual verification, it was observed that an entry for the emotion-bearing word '*bullshit*' is present in SentiWordNet while Inquirer and Taboada resource do not have them. This shows that the coverage of the lexical resource affects the performance of a system and may introduce errors.

4.2.5 Absence of Named Entity Recognition

Tweet: *@user I don't think I need to guess, but ok, close encounters of the third kind? Lol*
Entity: Close encounters of the third kind
Label by C-Feel-It: Positive

The words comprising the name of the film '*Close encounters of the third kind*' are also looked up. Inability to identify the named entity leads the system into this trap.

4.2.6 Requirement of World Knowledge

Tweet: *The soccer world cup boasts an audience twice that of the Summer Olympics.*
Label by C-Feel-It: Negative

To judge the opinion of this tweet, one requires an understanding of the fact that larger the audience, more favorable it is for a sports tournament. This world knowledge is important for a system that aims to handle tweets like these.

4.2.7 Mixed Emotion Tweets

Tweet: *oh but that last kiss tells me it's goodbye, just like nothing happened last night. but if i had one chance, i'd do it all over again*
Label by C-Feel-It: Positive

The tweet contains emotions of positive as well as negative variety and it would in fact be difficult for a human as well to identify the polarity. The mixed nature of the tweet leads to this error by the system.

4.2.8 Lack of Context

Tweet: *I'll have to say it's a tie between Little Women or To kill a Mockingbird*

Label by C-Feel-It: Negative
Label by human user: Positive

The tweet has a sentiment which will possibly be clear in the context of the conversation. Going by the tweet alone, while one understands that an comparative opinion is being expressed, it is not possible to tag it as positive or negative.

4.2.9 Concatenated Words

Tweet: *To Kill a Mockingbird is a #goodbook.*
Label by C-Feel-It: Negative

The tweet has a hashtag containing concatenated words '*goodbook*' which get overlooked as out-of-dictionary words and hence, are not used for sentiment prediction. The sentiment of '*good*' is not detected.

4.2.10 Interjections

Tweet: *Oooh. Apocalypse Now is on bluray now.*
Label by C-Feel-It: Objective
Label by human user: Positive

The extended interjection '*Oooh*' is an indicator of sentiment. Since it does not have a direct prior polarity, it is not present in any of the resources. However, this interjection is an important carrier of sentiment.

4.2.11 Comparatives

Tweet: *The more years I spend at Colbert Heights..the more disgusted I get by the people there. I'm soooo ready to graduate.*
Label by C-Feel-It: Positive
Label by human user: Negative

The comparatives in the sentence expressed by '*..more disgusted I get..*' have to be handled as a special case because '*more*' is an intensification of the negative sentiment expressed by the word '*disgusted*'.

5 Summary & Future Work

In this paper, we described a system which categorizes live tweets related to a keyword as positive, negative and objective based on the predictions of four sentiment-based resources. We also presented a qualitative evaluation of our system pointing out the areas of improvement for the current system.

A sentiment analyzer of this kind can be tuned to take inputs from different sources on the internet (for example, wall posts on facebook). In order to improve the quality of sentiment prediction, we propose two additions. Firstly, while we use simple heuristics to handle extensions of words in tweets, a deeper study is required to decipher the pragmatics involved. Secondly, a spam detection module that eliminates promotional tweets before performing sentiment detection may be added to the current system. Our goal with respect to this system is to deploy it for predicting share market values of firms based

on sentiment on social networks with respect to related entitites.

References

Go Alec, Huang Lei, and Bhayani Richa. 2009a. Twitter sentiment classification using distant supervision. Technical report, Standford University.

Go Alec, Bhayani Richa, Raghunathan Karthik, and Huang Lei. 2009b. May.

Andrea Esuli and Fabrizio Sebastiani. 2006. SentiWordNet: A publicly available lexical resource for opinion mining. In *Proceedings of LREC-06*, Genova, Italy.

Andreas M. Kaplan and Michael Haenlein. 2010. The early bird catches the news: Nine things you should know about micro-blogging. *Business Horizons*, 54(2):05 – 113.

Julie B. Lovins. 1968. Development of a Stemming Algorithm. June.

Bo Pang and Lillian Lee. 2004. A sentimental education: sentiment analysis using subjectivity summarization based on minimum cuts. In *Proceedings of the 42nd Annual Meeting on Association for Computational Linguistics*, ACL '04, Stroudsburg, PA, USA. Association for Computational Linguistics.

Bo Pang and Lillian Lee. 2005. Seeing stars: Exploiting class relationships for sentiment categorization with respect to rating scales. In *Proceedings of ACL-05*.

Vladimir Prelovac. 2010. Top social media sites. Web, May.

Philip J. Stone, Dexter C. Dunphy, Marshall S. Smith, and Daniel M. Ogilvie. 1966. *The General Inquirer: A Computer Approach to Content Analysis*. MIT Press.

Maite Taboada and Jack Grieve. 2004. Analyzing Appraisal Automatically. In *Proceedings of the AAAI Spring Symposium on Exploring Attitude and Affect in Text: Theories and Applications*, pages 158–161, Stanford, US.

2000. *Enriching the knowledge sources used in a maximum entropy part-of-speech tagger*, Stroudsburg, PA, USA. Association for Computational Linguistics.

Janyce Wiebe, Theresa Wilson, Rebecca Bruce, Matthew Bell, and Melanie Martin. 2004. Learning subjective language. *Computational Linguistics*, 30:277–308, September.